

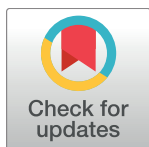
RESEARCH ARTICLE

Dengue Baidu Search Index data can improve the prediction of local dengue epidemic: A case study in Guangzhou, China

Zhihao Li¹, Tao Liu¹, Guanghu Zhu^{1,2}, Hualiang Lin¹, Yonghui Zhang³, Jianfeng He³, Aiping Deng³, Zhiqiang Peng³, Jianpeng Xiao¹, Shannon Rutherford⁴, Runsheng Xie¹, Weilin Zeng¹, Xing Li¹, Wenjun Ma^{1*}

1 Guangdong Provincial Institute of Public Health, Guangdong Provincial Center for Disease Control and Prevention, Guangzhou, China, **2** School of Mathematics and Computing Science, Guilin University of Electronic Technology, Guilin, China, **3** Guangdong Provincial Center for Disease Control and Prevention, Guangzhou, China, **4** Centre for Environment and Population Health, Griffith University, Brisbane, Australia

* mawj@gdiph.org.cn



OPEN ACCESS

Citation: Li Z, Liu T, Zhu G, Lin H, Zhang Y, He J, et al. (2017) Dengue Baidu Search Index data can improve the prediction of local dengue epidemic: A case study in Guangzhou, China. *PLoS Negl Trop Dis* 11(3): e0005354. doi:10.1371/journal.pntd.0005354

Editor: Benjamin Althouse, Institute for Disease Modeling, UNITED STATES

Received: August 11, 2016

Accepted: January 24, 2017

Published: March 6, 2017

Copyright: © 2017 Li et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This study was supported by Guangdong Provincial Science and Technology Project Fundings (NO.2013A020229005; NO.2014A040401041) and the National Natural Science Foundation of China (11661026). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Background

Dengue fever (DF) in Guangzhou, Guangdong province in China is an important public health issue. The problem was highlighted in 2014 by a large, unprecedented outbreak. In order to respond in a more timely manner and hence better control such potential outbreaks in the future, this study develops an early warning model that integrates internet-based query data into traditional surveillance data.

Methodology and principal findings

A Dengue Baidu Search Index (DBSI) was collected from the Baidu website for developing a predictive model of dengue fever in combination with meteorological and demographic factors. Generalized additive models (GAM) with or without DBSI were established. The generalized cross validation (GCV) score and deviance explained indexes, intraclass correlation coefficient (ICC) and root mean squared error (RMSE), were respectively applied to measure the fitness and the prediction capability of the models. Our results show that the DBSI with one-week lag has a positive linear relationship with the local DF occurrence, and the model with DBSI (ICC:0.94 and RMSE:59.86) has a better prediction capability than the model without DBSI (ICC:0.72 and RMSE:203.29).

Conclusions

Our study suggests that a DSBI combined with traditional disease surveillance and meteorological data can improve the dengue early warning system in Guangzhou.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Dengue fever is an important public health problem in China, and its importance was highlighted by an unprecedented outbreak in Guangdong province in 2014. Several previous studies have found that prediction models based on internet-based data have advantages in the timely detection of dengue epidemics. In this study, we employed the Dengue Baidu Search Index (DBSI) to explore whether internet-based query data can help improve disease prediction. We found that the dengue early warning system combining DBSI with traditional surveillance and meteorological data improved the prediction capability in Guangzhou, which suggests that utilizing big data from internet search engines can provide valuable supplementary data to traditional surveillance systems particularly for developing dengue early warning systems.

Introduction

Dengue fever (DF) is currently endemic in more than 100 countries, mainly in southeast Asia, the western Pacific islands and the Americas, with approximately 3.9 billion individuals at risk [1]. The annual number of infections is estimated at 390 million globally [2], making it one of the most significant vector-borne viral diseases.

In China, the first outbreak of DF was reported in Guangdong province in 1978 [3]. Since then DF cases have been reported in 26 provinces of China [4]. Guangdong province is the most affected areas in mainland China. In 2014, this province experienced a large outbreak resulting in 45,224 DF cases [5,6]. Since there is no specific treatment for DF and vector control remains the most effective way to prevent and control it [2]. Early warning systems are considered as one of the prerequisites for adequate preparedness and response to DF epidemics [7]. Several previous studies have reported meteorological factors that were associated with DF outbreaks through early warning models [8–12]. Among various meteorological factors, temperature and rainfall contribute the most to dengue epidemics [13]. In Singapore, Yien et al. [12] developed a weather-based dengue-forecasting model that allows warning 16 weeks in advance of dengue epidemics with high sensitivity and specificity. However, the dengue epidemics in Guangzhou are generally characterized by low level epidemic caused by imported cases, followed by a sudden and rapid transmission [14]. They have varied greatly in size from year to year [4], which poses a different challenge for prediction than in the more stable and endemic regions. Although a study conducted by Sang *et al.* [15] attempted to develop a model based on imported cases, minimum temperature and precipitation to predict the dengue incidence in Guangzhou, DF forecasting systems still face many difficulties due to the complexity of factors influencing DF outbreaks [16].

Over the past decade the increasing number of internet users around the world has provided new sources of data potentially useful for disease surveillance. This is increasingly being recognized as an opportunity to improve traditional disease surveillance systems [17]. For example, a study reported that using Google Flu Trends (GFT) could improve the prediction of influenza trends two weeks ahead of Centers for Disease Control and Prevention (CDC) reports in the US between 2003 and 2007 [18]. Several other studies using Google, Yahoo and other search data have been conducted worldwide to predict disease trends [19–24]. However, use of GFT data is not without its problems. For example, studies found that the surveillance data did not correspond with estimates provided by the GFT model in the US during the 2009 pandemic and the 2012/2013 epidemic season [25–27]. The reasons may be related to the proportion of the population who used the internet to obtain health-related information [17],

algorithm dynamics affecting Google's search algorithm [28] and media bias [29]. Therefore, researchers believe that internet search data is a good supplement to, rather than a substitute for, traditional disease surveillance data [28].

In China, Baidu is the most popular search engine, and approximately 86.7% of internet users prefer it [30]. Some recent studies have explored the potential of using Baidu search queries to predict diseases such as influenza [31] and erythromelalgia [32]. However, there has been no similar study in utilizing such data for DF prediction in China. Therefore, the aim of this study is to examine whether an early warning model utilizing internet-based dengue query data can improve DF prediction.

Materials and methods

Study setting

Guangzhou, the capital city of Guangdong province, is the third most populous city in China. At the end of 2014 the population in Guangzhou was 13.1 million [33]. This city is the center of transportation, finance, industry and trade in southern China and has a large exchange in business and tourism with southeast Asia, Africa and the Indian subcontinent. It has 12 districts with an area of 7473 km² and a typical subtropical monsoon climate, with an annual mean temperature of 22°C.

Data collection

DF has been a legally notifiable communicable disease in China since 1989. Weekly DF cases in Guangzhou during the period from January 1st, 2011 to December 31st, 2014 were retrieved from China Notifiable Infectious Disease Report System (NNIDRIS). DF cases before October 2014 were diagnosed according to the China National Diagnostic Criteria for dengue fever (WS216-2008) [34], and cases after October 2014 were diagnosed according to the new version of the China National Diagnostic Criteria for dengue fever (2014 version) enacted by the National Health and Family Planning Commission (<http://www.nhfpc.gov.cn>).

A climate dataset was obtained from the China Meteorological Data Sharing Service System (<http://cdc.nmic.cn/home.do>). It included weekly average minimum temperature (°C) and cumulative rainfall (mm) from 2011–2014. The population data was collected from the Guangzhou Statistical Yearbook.

The Baidu index database (<http://index.baidu.com>) contains search volumes for numerous terms entered by Baidu users since January 2011. The Baidu search query data are available as daily counts at the city, province and country level. We transformed the data to weekly counts for the analysis for consistency with other time series data.

As different terms have different search volumes and can therefore produce diverse models, term selection is the critical issue in internet search data-based surveillance. However, there are no criteria in practice [32,35,36]. Previous studies generally chose the nomenclature, clinical signs and symptoms of target diseases as the main terms [23,24,32]. Related terms were obtained from a Chinese website (<http://tool.chinaz.com/baidu/words.aspx>). Terms suggested by the website not only include recommendations from Baidu, but also from blogs, portal websites and online reports using semantic correlation analysis [31]. Upon typing in six primary terms, we obtained a total of 32 related search terms. More terms do not necessarily lead to a better result since some recommended terms are not closely related to DF occurrence, which could reduce the detective ability of the surveillance system [32]. Hence, we filtered terms following two steps. First, we eliminated the terms irrelevant to DF and those with a search volume of zero during the study period, and after these 26 keywords remained (S1 Table). Second, Spearman's rank correlation coefficients (ρ) were then calculated between weekly DF and

search volumes. We excluded the words with correlation coefficients smaller than 0.4 (S2 Table). Weights of terms were defined by the value of the correlation coefficient. The weights calculation and Dengue Baidu Search Index (DBSI) composition formulae are as follows:

$$weight_i = \frac{\rho_i}{\sum_1^n \rho_i}$$

$$Dengue\ Baidu\ Search\ Index = \sum_1^n weight_i term_i$$

Where n is the number of terms, $term_i$ and $weight_i$ represent the i^{th} term and the weight of it.

Statistical methods

First, cross-correlation analysis was carried out to identify the correlation between DF occurrence with imported cases, minimum temperature, cumulative rainfall and DBSI with 1 to 16 weeks' lag. Second, generalized additive models (GAM) were applied to fit the relationships between the variables and local DF cases. Because the variables with different time lags are highly correlated with each other, only those with maximal correlation coefficient were used to construct the model [15]. We used a cubic spline function for these variables to consider the non-linear association between factors and DF occurrence. In this study, a quasi-Poisson model was applied to allow for over-dispersion of the data. Model selection was based on the lowest generalized cross validation (GCV) scores.

In order to examine whether internet-based dengue query data could improve the prediction, we fit two models (with and without DBSI) and compared the relative parameters. Model (1) (without DBSI) and model (2) (with DBSI) are as follows:

$$\begin{aligned} \text{Log}(u_t) = & \beta_0 + s(\text{Tmin}_{t-e}, df) + s(\text{R}_{t-b}, df) + s(\text{Imp}_{t-c}, df) + s(\text{Local}_{t-1}, df) + \text{year} \\ & + s(\text{week}, df) + \text{offset}(\text{pop}) \end{aligned} \quad (1)$$

$$\begin{aligned} \text{Log}(u_t) = & \beta_0 + s(\text{DBSI}_{t-d}, df) + s(\text{Tmin}_{t-e}, df) + s(\text{R}_{t-b}, df) + s(\text{Imp}_{t-c}, df) + s(\text{Local}_{t-1}, df) \\ & + \text{year} + s(\text{week}, df) + \text{offset}(\text{pop}) \end{aligned} \quad (2)$$

where u_t represents the predicted mean DF cases during week t ; $s(\text{Tmin}_{t-e}, df)$ denotes the cubic spline of minimum temperature in the previous e weeks with corresponding df ; $s(\text{R}_{t-b}, df)$ represents the cubic spline of cumulative rainfall in the previous b weeks with corresponding df ; $s(\text{Imp}_{t-c}, df)$ represents the cubic spline of imported cases in the previous c weeks with corresponding df ; $s(\text{Local}_{t-1}, df)$ is the autoregressive term for local DF cases in the previous week with corresponding df ; $s(\text{DBSI}_{t-d}, df)$ denotes the cubic spline of DBSI in the previous d weeks with corresponding df ; year is used to control long-term trend, and $s(\text{week}, df)$ denotes the cubic spline of week with corresponding df that is used to control the seasonality; and $\text{offset}(\text{pop})$ accounts for population in Guangzhou during this period [12].

The df for each variable was determined according to the GCV principles and deviance explained (%) [15]. Lower GCV and higher deviance explained value indicate a better fit of the model. Finally, we chose df for week variable were 4, and other included variables were 3[37]; moreover, the sensitivity of the trend was tested by setting df to be 2, 3 or 4.

The disease dataset was also divided into two subsets: the first, from the 1st week of 2011 to the 44th week of 2014 was used for model construction, and the other, from the 45th to the 52nd week of 2014 for external validity assessment. We used the F test to compare the fit of models (with or without DBSI). Moreover, intraclass correlation coefficient (ICC) and root mean

squared error (RMSE) were applied to verify the consistency between the actual and predicted data [38,39].

Finally, we employed a Leave-One-Out Cross-Validation (LOOCV) method to validate the results of model (1) and model (2). LOOCV is a k-fold cross-validation method [40], and here the total dataset was divided into (n-1) subsets, where n is the number of weeks from the 1st week of 2011 to the 44th week of 2014. In each subset, a single week's data was removed, and the weekly number of dengue cases was predicted. Then we employed the ICC as a metric to test the correlations between predicted and observed cases.

All the analyses were performed using the “mgcv” library in R 3.2.2 [41] with a significance level of $P < 0.05$.

Results

During 2011–2014, a total of 38,860 DF cases were reported in Guangzhou city, with 116 (0.3%) imported DF cases and 38,744 (99.7%) local DF cases. A summary of meteorological variables, DBSI and DF cases are presented in Table 1. There was an average of 186.3 local DF cases and 0.6 imported DF cases every week over the study period. The mean values of the weekly DBSI, minimum temperature and cumulative rainfall were 80.8, 19.0°C and 34.3mm, respectively. Fig 1 shows the time series of weekly meteorological variables, DBSI, and local and imported DF cases. Both a large DF outbreak and the highest weekly DBSI during the study period occurred in 2014. Weekly minimum temperature and cumulative rainfall showed an obvious seasonal pattern, peaking from June to August.

The results of the cross-correlation of weekly local DF case numbers and prediction variables are shown in S3 Table. We found that minimum temperature in the previous 9 weeks, cumulative rainfall in the previous 12 weeks and imported cases in the previous 5 weeks have the highest correlation with local DF. Hence these variables were included in our model. Fig 2 shows the dose-response relationship between local DF cases and imported cases in the previous 5 weeks, minimum temperature in the previous 9 weeks, cumulative rainfall in the previous 12 weeks and DBSI in the previous week. Minimum temperature, cumulative rainfall and imported DF cases were non-linearly associated with the local DF cases. For cumulative rainfall, the risk of DF incidence increases with the increment of rainfall at first, peaking at 149mm, followed by a significant decrease. DBSI in the previous week had a positively linear relationship with the local DF.

Fig 3 shows that both model (1) and model (2) fit the DF cases reasonably well during the training process. Our results indicate that the fit of model (1) and model (2) were both found to be significant ($F = 10.46$, $P < 0.001$). The value of model with the DBSI (GCV: 7.62 and Deviance explained: 99.23%) fit better than the model without DBSI (GCV: 18.41 and Deviance explained: 94.53%). Moreover, the effects of climate, imported cases and DBSI were found to be significant at the 0.05 level (S4 Table). The one-week ahead predictions of dengue outbreaks that occurred from the 45th week to the 52nd week of 2014 for both models are shown in Fig 4.

Table 1. Weekly meteorological parameters, DBSI, and local and imported DF cases in Guangzhou from 2011–2014.

Variables	Min	Mean	Median	Max	SD
Minimum temperature(°C)	5.4	19.0	20.3	26.6	6.3
Cumulative rainfall (mm)	0.0	34.3	16.5	324.0	46.3
DBSI	4.9	80.8	21.8	1890.3	234.8
Weekly local DF count	0.0	186.3	0.0	8200.0	929.3
Weekly imported DF count	0.0	0.6	0.0	6.0	1.0

doi:10.1371/journal.pntd.0005354.t001

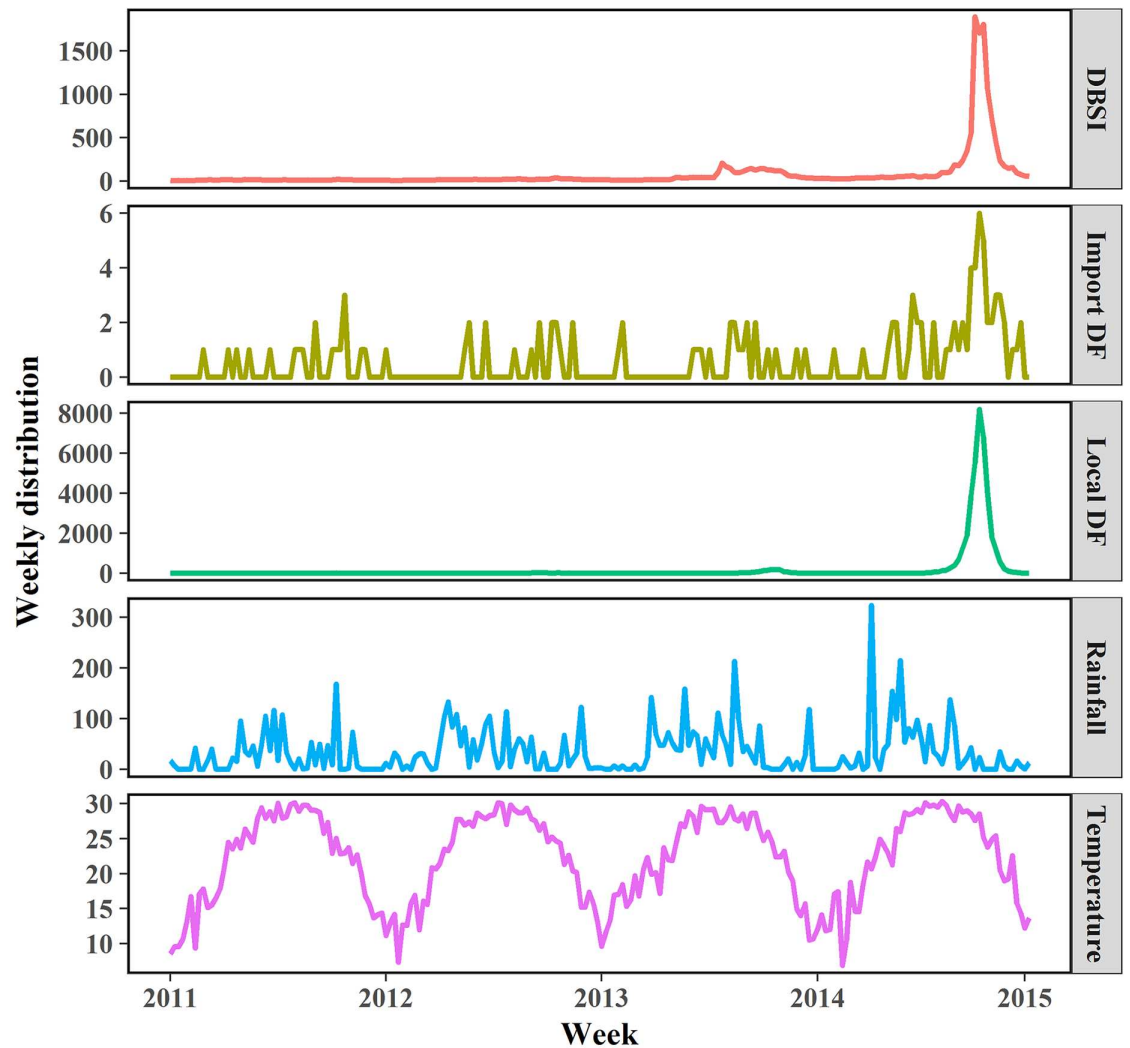


Fig 1. The time series of weekly rainfall, minimum temperature, DBSI, local dengue and imported dengue counts from 2011–2014 in Guangzhou.

doi:10.1371/journal.pntd.0005354.g001

Model (2) gives a better prediction of DF cases (ICC:0.94 and RMSE:59.86) than model (1) (ICC:0.72 and RMSE:203.29).

The results of sensitivity analyses show that the GCVs were respectively the lowest when the *dfs* of weekly minimum temperature, and cumulative rainfall in model (1) and DBSI in model (2) were set to 3, which justified the *df* selection in our models (S5 Table and S6 Table). In addition, the results of LOOCV also showed that the performance of model (2) was better than model (1) (S7 Table).

Discussion

DF has become an increasingly important public health concern in Guangzhou, China in recent years, and in 2014 the number of DF cases represented the highest peak in the past 25 years [4]. A recent study suggested that urbanization, climate change, international trade and population movement were important factors that influenced this re-emergence of dengue in

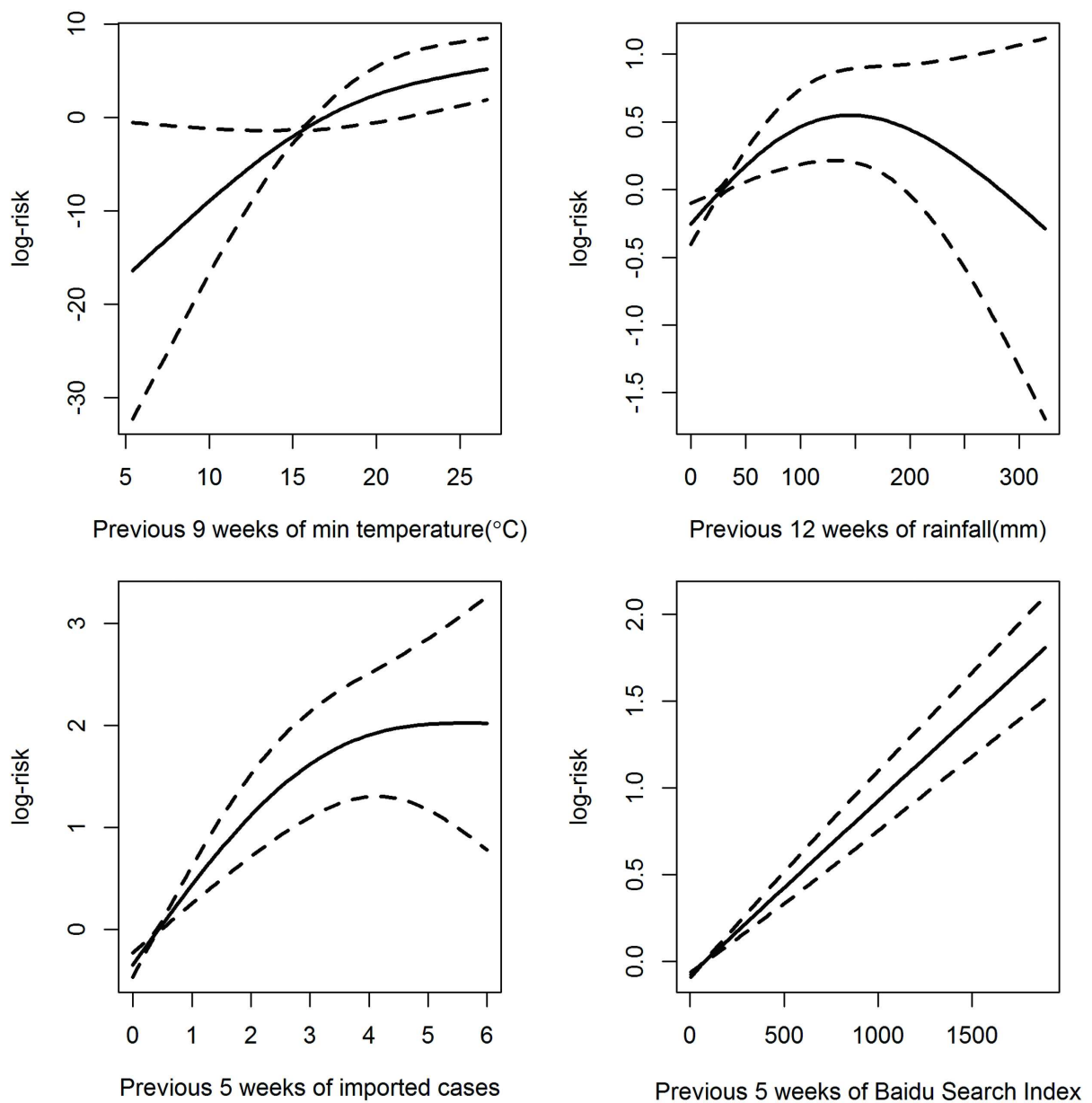


Fig 2. The relationship between the minimum temperature at 9 weeks' lag, imported DF at 5 weeks' lag, cumulative rainfall at 12 weeks lag and DBSI for the previous week with local DF cases. Note: Solid lines represent logarithmic relative risks of DF and dotted lines represent the upper and lower limits of 95% confidence intervals

doi:10.1371/journal.pntd.0005354.g002

Guangzhou [5]. In order to improve early and rapid response to dengue outbreaks in Guangzhou, we combined dengue internet-based data (DBSI) with imported cases, temperature and rainfall to develop an early warning model. We found that inclusion of DBSI can improve the prediction of the base model reliant on traditional disease surveillance data. The results provide a new approach to developing a dengue early warning system in Guangzhou.

Many previous studies reported that climatic factors influenced DF transmission by directly or indirectly affecting each stage in the life cycle of the mosquito and the disease transmission [42,43]. In this study, we found that DF was positively correlated with average weekly minimum

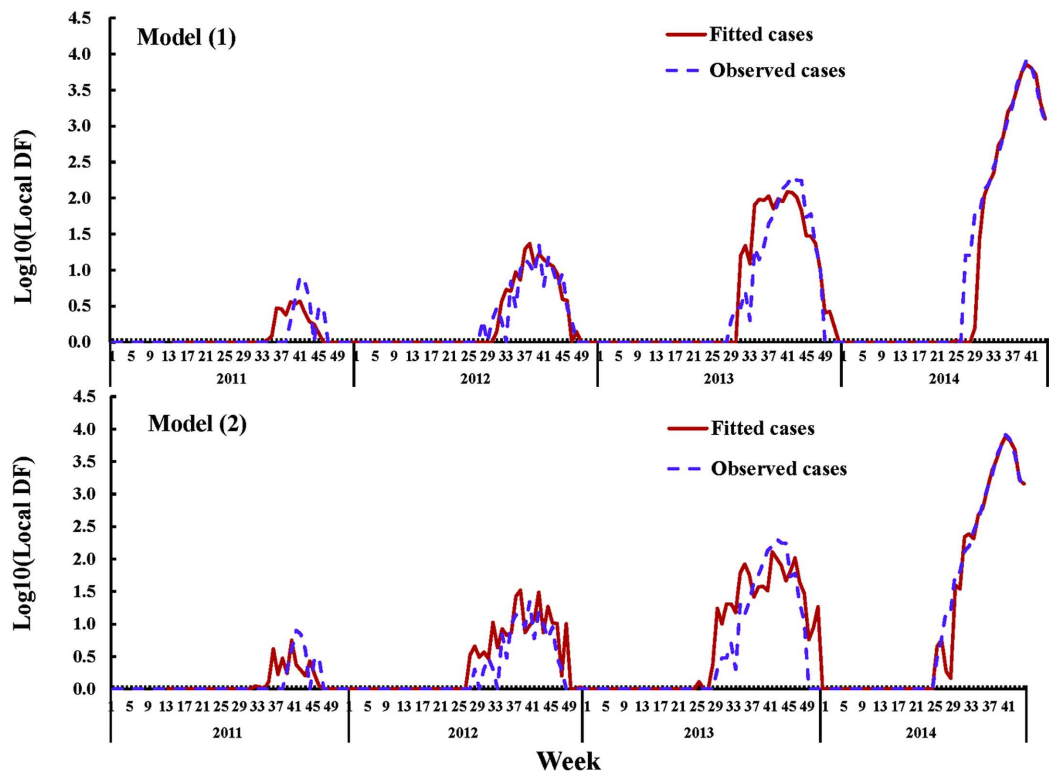


Fig 3. Weekly observed and fitted local DF cases using two different models from 1st week of 2011 to the 44th week of 2014. Note: The dotted line represents the reported dengue cases and the solid lines represent the cases fitted by the respective models.

doi:10.1371/journal.pntd.0005354.g003

temperature at a lag of 9 weeks. This finding is generally in agreement with previous studies that indicate the crucial role of temperature in dengue transmission [44,45]. Possible reasons for this association with temperature are that higher temperature can reduce both mosquito maturity and reproduction time in favor of producing more mosquitoes in a shorter time [46]. We also found that rainfall has a nonlinear relationship with DF with a threshold of 149mm. This is consistent with several other studies that found that rainfall influenced vector abundance in subsequent weeks by creating more breeding habitats for mosquitoes [47]. On the other hand, it is also likely that heavy rain can destroy existing mosquito breeding sites and affect the maturation of mosquito eggs or larvae [48]. DF is not regarded as endemic in Guangzhou, and previous outbreaks were caused by imported cases [14]. Our study indicated that imported DF cases in the previous 5 weeks had a large impact on the local DF case numbers. The time delay could be the period of the life cycle of dengue transmission.

To the best of our knowledge, this study is the first one to investigate the relationship between DBSI and DF cases in China. We found that DBSI in the previous week had a positive linear relationship with reported DF cases, implying that internet-based search behavior may be a useful predictor of DF incidence. This is consistent with previous studies that investigated the relationship between Google Dengue Trends (GDT) and DF cases [23,24,49]. In one study in Singapore and Bangkok, Althouse et al. demonstrated that the internet search terms could successfully predict incidence and periods of large incidence of dengue with high accuracy. Their model using Google search data had an $r^2 = 0.948$ and 0.943 for Singapore and Bangkok

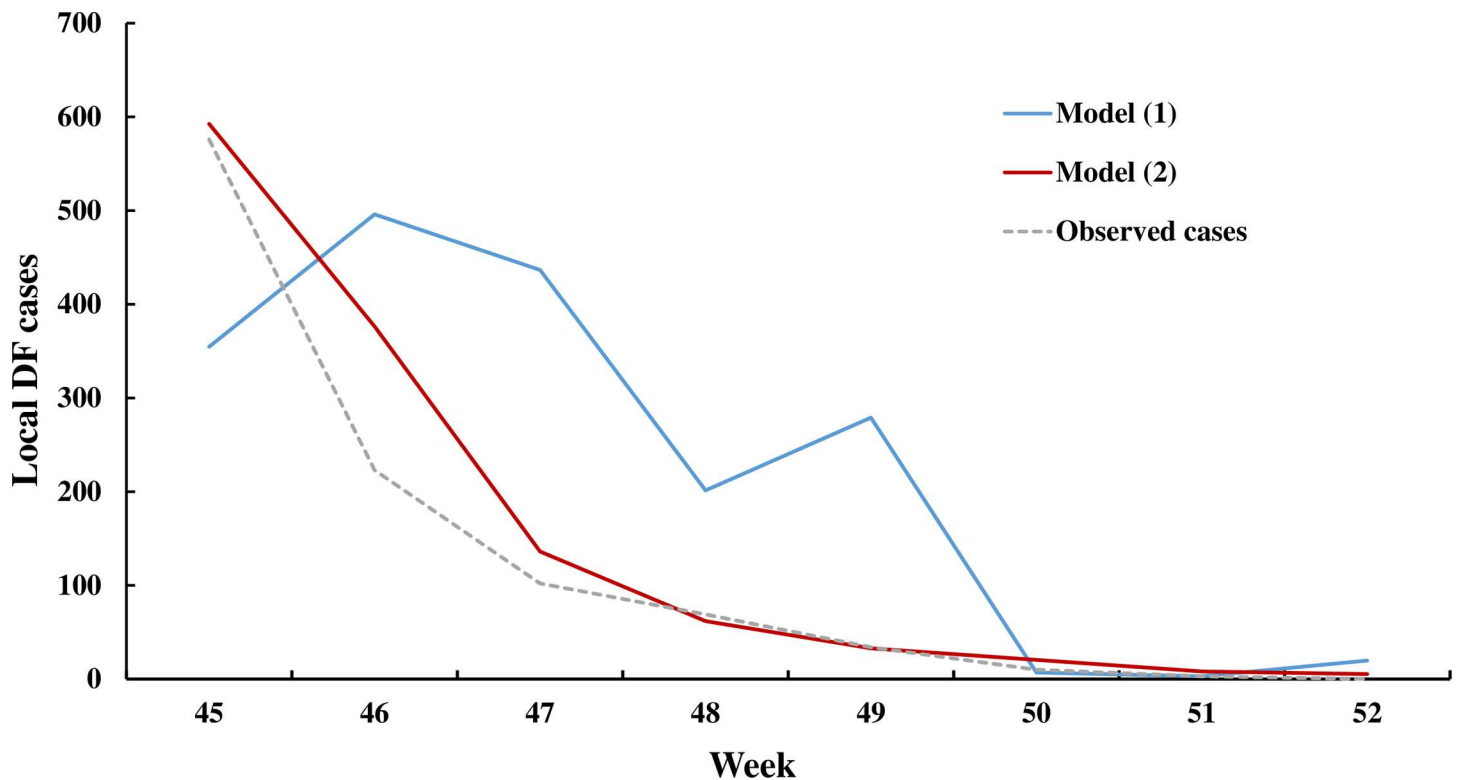


Fig 4. Comparison of weekly-observed DF cases from the 45th week to the 52nd week of 2014 with out-of-sample one-week ahead predictions. Note: The dotted line represents observed dengue cases and the solid lines show the cases predicted by the fitted models.

doi:10.1371/journal.pntd.0005354.g004

[23]. Chan et al. also observed in five countries that the models built on the fraction of Google search volume for dengue-related queries were able to adequately estimate true dengue activity, and the correlation between values predicted by models and the surveillance data was generally quite high, ranging from 0.82 to 0.99 [24].

As we mentioned in the introduction, the GFT firstly provided us an excellent example in 2003–2007 to illustrate the contribution of internet search data on the prediction of infectious diseases [18]. However, the GFT failed to successfully predict the seasonal and pandemic influenza in the USA during the 2012/2013 season [25]. It has been debated that the internet-based query might misrepresent the epidemic curve in practice [25,29]. Some researchers analyzed the reasons for this failure in the GFT model suggesting that the internet-based query system can be used as a supplement to, but not a substitute for, the traditional data collection and analysis [17,50]. Moreover, Gluskin et al. also demonstrated in Mexico that the model using GDT data in combination with relevant covariates (maximum temperature, logged precipitation) can significantly improve dengue prediction [49]. In our study, similar result was also found that the model including DBSI variable had a better performance than model without it. Collectively, these results indicate that integrating internet-based dengue query data into traditional disease surveillance can improve dengue prediction, providing us with a new approach for establishing an almost real-time early warning system. In this big data era when internet-based data are easily available and collected in almost real-time [51], its use as a supplement to traditional disease surveillance provides important progress towards establishing reliable early warning models allowing for more efficient and rapid control of infectious diseases.

We validated our model by comparing the predictive results with the surveillance dengue data in the last 8 weeks of study periods, and the results show good performance of the model.

However, it has been suggested that the results of models using internet search queries need to be further validated by more advanced studies to control the relevant covariates (such as media basis, socio-economic and demographic factors) [50].

Some limitations of our study should be mentioned. First, the guidelines of dengue diagnosis and treatment were different before and after October 11th, 2014 in China. For example, a dengue virus NS1 antigen test was added to the new version as an important criterion, which might lead to some bias to our results. However, the influences of changing diagnosis guidelines on our results are limited, because only dengue cases in the last one and half months in 2014 were diagnosed by the new guidelines. Second, the study developed the prediction model using only a 4-year period of time-series data based on weekly data, and could only be validated for an 8-week period. It is advisable to use long-term time series data in model fitting in the future. Third, this study does not examine other potential confounding factors that may be associated with dengue incidence, such as environmental, socio-economic and demographic factors [52]. In addition, it has been suggested that internet searching behavior is susceptible to the impact of media reports [23,53], and we did not implement any measures to control for this. Studies could be conducted in the future to investigate how users interact with internet search sources for providing valuable information on potential biases and suggest mechanisms for improving the robustness of surveillance systems based on internet search queries.

Conclusions

The present study suggests that the Dengue Baidu Search Index provides useful data for early prediction of a dengue outbreak. Such improvements in prediction and hence early warning are very important for improving prevention and control of dengue epidemics in the future.

Supporting information

S1 Table. Search terms from Baidu in Chinese and English.

(DOCX)

S2 Table. Dengue related Baidu search terms that were finally selected.

(DOCX)

S3 Table. Cross-correlation coefficients for local DF cases in Guangzhou and four predicting variables.

(DOCX)

S4 Table. Effective degrees of freedom of the smooth function terms in Model (2).

(DOCX)

S5 Table. Sensitivity analyses on the effects of df on GCVs in model (1).

(DOCX)

S6 Table. Sensitivity analyses on the effects of df on GCVs in model (2).

(DOCX)

S7 Table. The ICCs of model (1) and model (2) validated by the LOOCV method.

(DOCX)

Author Contributions

Conceptualization: WM YZ.

Data curation: JH AD ZP.

Formal analysis: JX RX WZ XL.

Methodology: TL GZ WM.

Resources: JH AD ZP.

Software: TL HL.

Supervision: WM HL TL.

Validation: WM.

Visualization: ZL WM HL TL.

Writing – original draft: ZL WM HL JH AD ZP SR.

Writing – review & editing: ZL WM HL TL.

References

1. Brady OJ, Gething PW, Bhatt S, Messina JP, Brownstein JS, et al. (2012) Refining the global spatial limits of dengue virus transmission by evidence-based consensus. *PLoS Negl Trop Dis* 6: e1760. doi: [10.1371/journal.pntd.0001760](https://doi.org/10.1371/journal.pntd.0001760) PMID: [22880140](https://pubmed.ncbi.nlm.nih.gov/22880140/)
2. Bhatt S, Gething PW, Brady OJ, Messina JP, Farlow AW, et al. (2013) The global distribution and burden of dengue. *Nature* 496: 504–507. doi: [10.1038/nature12060](https://doi.org/10.1038/nature12060) PMID: [23563266](https://pubmed.ncbi.nlm.nih.gov/23563266/)
3. Zhao HL, Luo QH, Shen G (1981) Epidemiology of the dengue outbreak in Shiwanzhen, Nanhai County, Guangdong Province. *Chin Med J*. 61: 466–469.
4. Lai S, Huang Z, Zhou H, Anders KL, Perkins TA, et al. (2015) The changing epidemiology of dengue in China, 1990–2014: a descriptive analysis of 25 years of nationwide surveillance data. *BMC Med* 13:1–12.
5. Xiao J-P, He J-F, Deng A-P, Lin H-L, Song T, et al. (2016) Characterizing a large outbreak of dengue fever in Guangdong Province, China. *Infectious Diseases of Poverty* 5: 1–8.
6. Lin H, Liu T, Song T, Lin L, Xiao J, et al. (2016) Community Involvement in Dengue Outbreak Control: An Integrated Rigorous Intervention Strategy. *PLoS Negl Trop Dis* 10: e0004919. doi: [10.1371/journal.pntd.0004919](https://doi.org/10.1371/journal.pntd.0004919) PMID: [27548481](https://pubmed.ncbi.nlm.nih.gov/27548481/)
7. World Health Organization. (2009). Special Programme for Research and Training in Tropical Disease. Dengue: guidelines for diagnosis, treatment, prevention and control. Available: http://whqlibdoc.who.int/publications/2009/9789241547871_eng.pdf.
8. Thai KT, Cazelles B, Nguyen NV, Vo LT, Boni MF, et al. (2010) Dengue dynamics in Binh Thuan province, southern Vietnam: periodicity, synchronicity and climate variability. *PLoS Negl Trop Dis* 4: e747. doi: [10.1371/journal.pntd.0000747](https://doi.org/10.1371/journal.pntd.0000747) PMID: [20644621](https://pubmed.ncbi.nlm.nih.gov/20644621/)
9. Banu S, Hu W, Guo Y, Hurst C, Tong S (2014) Projecting the impact of climate change on dengue transmission in Dhaka, Bangladesh. *Environment international* 63: 137–142. doi: [10.1016/j.envint.2013.11.002](https://doi.org/10.1016/j.envint.2013.11.002) PMID: [24291765](https://pubmed.ncbi.nlm.nih.gov/24291765/)
10. Sang S, Yin W, Bi P, Zhang H, Wang C, et al. (2014) Predicting local dengue transmission in Guangzhou, China, through the influence of imported cases, mosquito density and climate variability. *PLoS One* 9: e102755. doi: [10.1371/journal.pone.0102755](https://doi.org/10.1371/journal.pone.0102755) PMID: [25019967](https://pubmed.ncbi.nlm.nih.gov/25019967/)
11. Gharbi M, Quenel P, Gustave J, Cassadou S, La Ruche G, et al. (2011) Time series analysis of dengue incidence in Guadeloupe, French West Indies: forecasting models using climate variables as predictors. *BMC infectious diseases* 11: 1.
12. Hii YL, Zhu H, Ng N, Ng LC, Rocklöv J (2012) Forecast of dengue incidence using temperature and rainfall. *PLoS Negl Trop Dis* 6: e1908. doi: [10.1371/journal.pntd.0001908](https://doi.org/10.1371/journal.pntd.0001908) PMID: [23209852](https://pubmed.ncbi.nlm.nih.gov/23209852/)
13. Louis VR, Phalkey R, Horstick O, Ratanawong P, Wilder-Smith A, et al. (2014) Modeling tools for dengue risk mapping—a systematic review. *International journal of health geographics* 13: 1.
14. Sang S, Chen B, Wu H, Yang Z, Di B, et al. (2015) Dengue is still an imported disease in China: a case study in Guangzhou. *Infection, Genetics and Evolution* 32: 178–190. doi: [10.1016/j.meegid.2015.03.005](https://doi.org/10.1016/j.meegid.2015.03.005) PMID: [25772205](https://pubmed.ncbi.nlm.nih.gov/25772205/)
15. Sang S, Gu S, Bi P, Yang W, Yang Z, et al. (2015) Predicting unprecedented dengue outbreak using imported cases and climatic factors in Guangzhou, 2014. *PLoS Negl Trop Dis* 9: e0003808. doi: [10.1371/journal.pntd.0003808](https://doi.org/10.1371/journal.pntd.0003808) PMID: [26020627](https://pubmed.ncbi.nlm.nih.gov/26020627/)

16. Racloz V, Ramsey R, Tong S, Hu W (2012) Surveillance of dengue fever virus: a review of epidemiological models and early warning systems. *PLoS Negl Trop Dis* 6: e1648. doi: [10.1371/journal.pntd.0001648](https://doi.org/10.1371/journal.pntd.0001648) PMID: [22629476](https://pubmed.ncbi.nlm.nih.gov/22629476/)
17. Milinovich GJ, Williams GM, Clements AC, Hu W (2014) Internet-based surveillance systems for monitoring emerging infectious diseases. *The Lancet infectious diseases* 14: 160–168. doi: [10.1016/S1473-3099\(13\)70244-5](https://doi.org/10.1016/S1473-3099(13)70244-5) PMID: [24290841](https://pubmed.ncbi.nlm.nih.gov/24290841/)
18. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, et al. (2009) Detecting influenza epidemics using search engine query data. *Nature* 457: 1012–1014. doi: [10.1038/nature07634](https://doi.org/10.1038/nature07634) PMID: [19020500](https://pubmed.ncbi.nlm.nih.gov/19020500/)
19. Polgreen PM, Chen Y, Pennock DM, Nelson FD, Weinstein RA (2008) Using internet searches for influenza surveillance. *Clinical infectious diseases* 47: 1443–1448. doi: [10.1086/593098](https://doi.org/10.1086/593098) PMID: [18954267](https://pubmed.ncbi.nlm.nih.gov/18954267/)
20. Kang M, Zhong H, He J, Rutherford S, Yang F (2013) Using google trends for influenza surveillance in South China. *PloS one* 8: e55205. doi: [10.1371/journal.pone.0055205](https://doi.org/10.1371/journal.pone.0055205) PMID: [23372837](https://pubmed.ncbi.nlm.nih.gov/23372837/)
21. Dugas AF, Jalalpour M, Gel Y, Levin S, Torcaso F, et al. (2013) Influenza forecasting with Google flu trends. *PloS one* 8: e56176. doi: [10.1371/journal.pone.0056176](https://doi.org/10.1371/journal.pone.0056176) PMID: [23457520](https://pubmed.ncbi.nlm.nih.gov/23457520/)
22. Thompson L, Malik M, Gumel A, Strome T, Mahmud S (2014) Emergency department and ‘Google flu trends’ data as syndromic surveillance indicators for seasonal influenza. *Epidemiology and infection* 142: 2397–2405. doi: [10.1017/S0950268813003464](https://doi.org/10.1017/S0950268813003464) PMID: [24480399](https://pubmed.ncbi.nlm.nih.gov/24480399/)
23. Althouse BM, Ng YY, Cummings DA (2011) Prediction of dengue incidence using search query surveillance. *PLoS Negl Trop Dis* 5: e1258. doi: [10.1371/journal.pntd.0001258](https://doi.org/10.1371/journal.pntd.0001258) PMID: [21829744](https://pubmed.ncbi.nlm.nih.gov/21829744/)
24. Chan EH, Sahai V, Conrad C, Brownstein JS (2011) Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. *PLoS Negl Trop Dis* 5: e1206. doi: [10.1371/journal.pntd.0001206](https://doi.org/10.1371/journal.pntd.0001206) PMID: [21647308](https://pubmed.ncbi.nlm.nih.gov/21647308/)
25. Olson DR, Konty KJ, Paladini M, Viboud C, Simonsen L (2013) Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. *PLoS Comput Biol* 9: e1003256. doi: [10.1371/journal.pcbi.1003256](https://doi.org/10.1371/journal.pcbi.1003256) PMID: [24146603](https://pubmed.ncbi.nlm.nih.gov/24146603/)
26. Santillana M, Zhang DW, Althouse BM, Ayers JW (2014) What can digital disease detection learn from (an external revision to) Google Flu Trends? *American journal of preventive medicine* 47: 341–347. doi: [10.1016/j.amepre.2014.05.020](https://doi.org/10.1016/j.amepre.2014.05.020) PMID: [24997572](https://pubmed.ncbi.nlm.nih.gov/24997572/)
27. Achee NL, Gould F, Perkins TA, Reiner RC Jr, Morrison AC, et al. (2015) A critical assessment of vector control for dengue prevention. *PLoS Negl Trop Dis* 9: e0003655. doi: [10.1371/journal.pntd.0003655](https://doi.org/10.1371/journal.pntd.0003655) PMID: [25951103](https://pubmed.ncbi.nlm.nih.gov/25951103/)
28. Lazer D, Kennedy R, King G, Vespignani A (2014) The parable of Google flu: traps in big data analysis. *Science* 343: 1203–1205. doi: [10.1126/science.1248506](https://doi.org/10.1126/science.1248506) PMID: [24626916](https://pubmed.ncbi.nlm.nih.gov/24626916/)
29. Butler D (2013) When Google got flu wrong. *Nature* 494: 155. doi: [10.1038/494155a](https://doi.org/10.1038/494155a) PMID: [23407515](https://pubmed.ncbi.nlm.nih.gov/23407515/)
30. China Internet Network Information Center. 2013 Chinese search engine market research report. (2014) Available: <http://www.cnnic.net.cn/hlwfzjy/hlxzbg/ssbg/201401/P020140127366465515288.pdf>.
31. Yuan Q, Nsoesie EO, Lv B, Peng G, Chunara R, et al. (2013) Monitoring influenza epidemics in China with search query from Baidu. *PloS one* 8: e64323. doi: [10.1371/journal.pone.0064323](https://doi.org/10.1371/journal.pone.0064323) PMID: [23750192](https://pubmed.ncbi.nlm.nih.gov/23750192/)
32. Gu Y, Chen F, Liu T, Lv X, Shao Z, et al. (2015) Early detection of an epidemic erythromelalgia outbreak using Baidu search data. *Scientific reports* 5.
33. Statistics Bureau of Guangzhou Municipality. (2015) Guangzhou Economic and Social Development Statistics Bulletin. Available: http://gzdaily.dayoo.com/html/2015-03/22/content_2887547.html.
34. Jing Q-L, Yang Z-C, Luo L, Xiao X-C, Di B, et al. (2012) Emergence of dengue virus 4 genotype II in Guangzhou, China, 2010: survey and molecular epidemiology of one community outbreak. *BMC infectious diseases* 12: 1.
35. Jia-xing B, Bcn-fu L, Geng P, Na L. Gonorrhoea incidence forecasting research based on Baidu search data; 2013. *IEEE*. pp. 36–42.
36. Liu Y, Lv B, Peng G, Yuan Q. A preprocessing method of internet search data for prediction improvement: application to Chinese stock market; 2012. *ACM*. pp. 3.
37. Hii YL, Rocklöv J, Wall S, Ng LC, Tang CS, et al. (2012) Optimal lead time for dengue forecast. *PLoS Negl Trop Dis* 6: e1848. doi: [10.1371/journal.pntd.0001848](https://doi.org/10.1371/journal.pntd.0001848) PMID: [23110242](https://pubmed.ncbi.nlm.nih.gov/23110242/)
38. Dayama P, Kameshwaran S. Predicting the dengue incidence in Singapore using univariate time series models; 2013. *American Medical Informatics Association*. pp. 285.

39. Santillana M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, et al. (2015) Combining search, social media, and traditional data sources to improve influenza surveillance. *PLoS Comput Biol* 11: e1004513. doi: [10.1371/journal.pcbi.1004513](https://doi.org/10.1371/journal.pcbi.1004513) PMID: [26513245](https://pubmed.ncbi.nlm.nih.gov/26513245/)
40. Kohavi R. (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection; pp. 1137–1145.
41. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
42. Halstead SB (2008) Dengue virus-mosquito interactions. *Annu Rev Entomol* 53: 273–291. doi: [10.1146/annurev.ento.53.103106.093326](https://doi.org/10.1146/annurev.ento.53.103106.093326) PMID: [17803458](https://pubmed.ncbi.nlm.nih.gov/17803458/)
43. Yang H, Macoris M, Galvani K, Andrighetti M, Wanderley D (2009) Assessing the effects of temperature on dengue transmission. *Epidemiology and Infection* 137: 1179–1187. doi: [10.1017/S0950268809002052](https://doi.org/10.1017/S0950268809002052) PMID: [19192323](https://pubmed.ncbi.nlm.nih.gov/19192323/)
44. Qi X, Wang Y, Li Y, Meng Y, Chen Q, et al. (2015) The Effects of Socioeconomic and Environmental Factors on the Incidence of Dengue Fever in the Pearl River Delta, China, 2013. *PLoS Negl Trop Dis* 9: e0004159. doi: [10.1371/journal.pntd.0004159](https://doi.org/10.1371/journal.pntd.0004159) PMID: [26506616](https://pubmed.ncbi.nlm.nih.gov/26506616/)
45. SHEN JC, Lei L, Li L, JING QL, OU CQ, et al. (2015) The impacts of mosquito density and meteorological factors on dengue fever epidemics in Guangzhou, China, 2006–2014: a time-series analysis. *Bio-medical and Environmental Sciences* 28: 321–329. doi: [10.3967/bes2015.046](https://doi.org/10.3967/bes2015.046) PMID: [26055559](https://pubmed.ncbi.nlm.nih.gov/26055559/)
46. Farjana T, Tuno N, Higa Y (2012) Effects of temperature and diet on development and interspecies competition in *Aedes aegypti* and *Aedes albopictus*. *Medical and veterinary entomology* 26: 210–217. doi: [10.1111/j.1365-2915.2011.00971.x](https://doi.org/10.1111/j.1365-2915.2011.00971.x) PMID: [21781139](https://pubmed.ncbi.nlm.nih.gov/21781139/)
47. Karl S, Halder N, Kelso JK, Ritchie SA, Milne GJ (2014) A spatial simulation model for dengue virus infection in urban areas. *BMC infectious diseases* 14: 1.
48. Dieng H, Rahman GS, Hassan AA, Salmah MC, Satho T, et al. (2012) The effects of simulated rainfall on immature population dynamics of *Aedes albopictus* and female oviposition. *International journal of biometeorology* 56: 113–120. doi: [10.1007/s00484-011-0402-0](https://doi.org/10.1007/s00484-011-0402-0) PMID: [21267602](https://pubmed.ncbi.nlm.nih.gov/21267602/)
49. Gluskin RT, Johansson MA, Santillana M, Brownstein JS (2014) Evaluation of Internet-based dengue query data: Google Dengue Trends. *PLoS Negl Trop Dis* 8: e2713. doi: [10.1371/journal.pntd.0002713](https://doi.org/10.1371/journal.pntd.0002713) PMID: [24587465](https://pubmed.ncbi.nlm.nih.gov/24587465/)
50. Althouse BM, Scarpino SV, Meyers LA, Ayers JW, Bargsten M, et al. (2015) Enhancing disease surveillance with novel data streams: challenges and opportunities. *EPJ Data Science* 4: 1.
51. Madoff LC, Fisman DN, Kass-Hout T (2011) A new approach to monitoring dengue activity. *PLoS Negl Trop Dis* 5: e1215. doi: [10.1371/journal.pntd.0001215](https://doi.org/10.1371/journal.pntd.0001215) PMID: [21647309](https://pubmed.ncbi.nlm.nih.gov/21647309/)
52. Hu W, Clements A, Williams G, Tong S, Mengersen K (2012) Spatial patterns and socioecological drivers of dengue fever transmission in Queensland, Australia. *Environmental health perspectives* 120: 260. doi: [10.1289/ehp.1003270](https://doi.org/10.1289/ehp.1003270) PMID: [22015625](https://pubmed.ncbi.nlm.nih.gov/22015625/)
53. Valdivia A, Lopez-Alcalde J, Vicente M, Pichiule M, Ruiz M, et al. (2010) Monitoring influenza activity in Europe with Google Flu Trends: comparison with the findings of sentinel physician networks—results for 2009–10. *Eurosurveillance* 15: 2–7.